RNN-Recurrent Neural Network

What is RNN ?

It is special type of neural network

It is sequential model and is used to work on sequential data

Sequential data - data where sequence matters

1. Text
2. Time series data

Why not ANN ?

In ANN we feed all the information at once therefore important sequential information is lost

ANN do not perform well on sequential data

Textual data can be of different sizes and we can’t vary the input size in ANN

RNN is being heavily used in natural language processing

Applications of RNN

-

- Sentiment analysis
- Sentence completion
- Image caption generation
- Google translate

Data for RNN

RNN requires input data in very specific format

Let’s say we have following dataset -

Reviews	Sentiment
Movie was good	1
Movie was bad	0
Movie was not good	o

We feed the data to RNN in the following format -

(timesteps , input features)

Our vocabulary has 5 distinct words

movie=[1,0,0,0,0] was=[0,1,0,0,0] good=[0,0,1,0,0] not=[0,0,0,0,1] bad=[0,0,0,1,0]

Review 1 i.e. movie was good can be represented as -

[[1,0,0,0,0],[0,1,0,0,0],[0,0,1,0,0]]

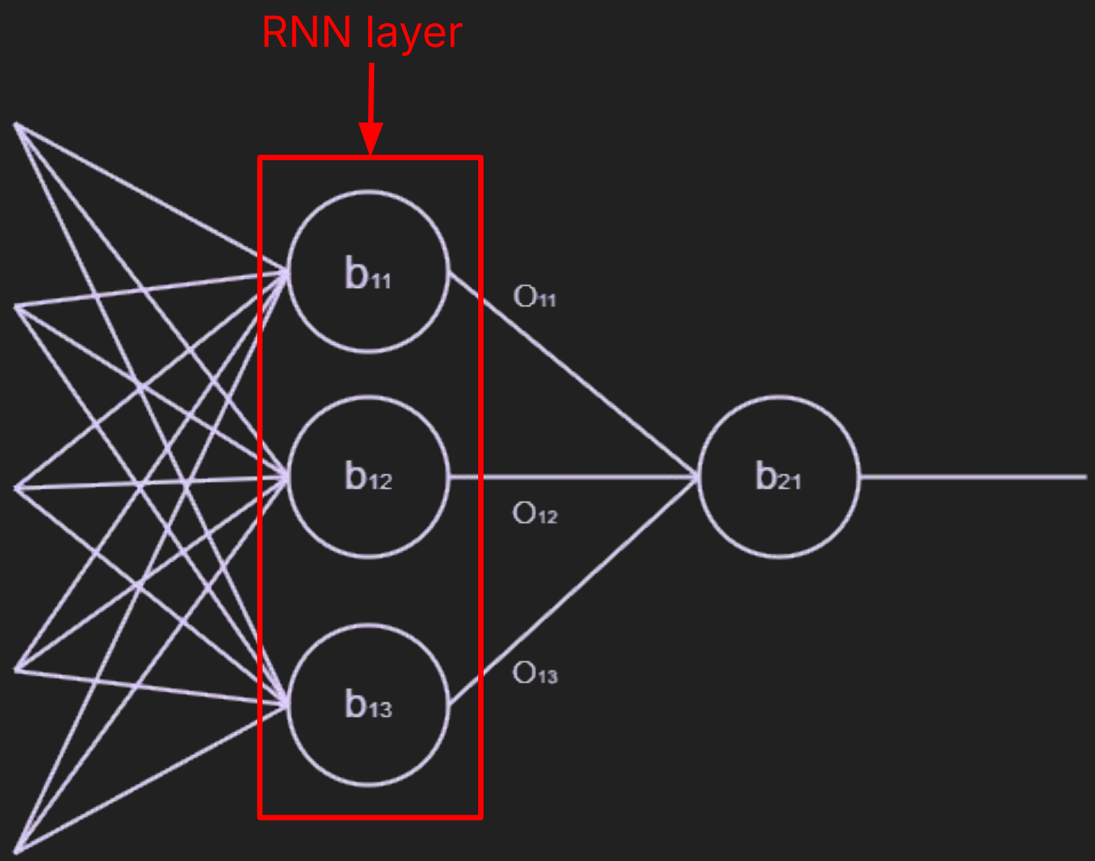
In RNN we feed data word by word

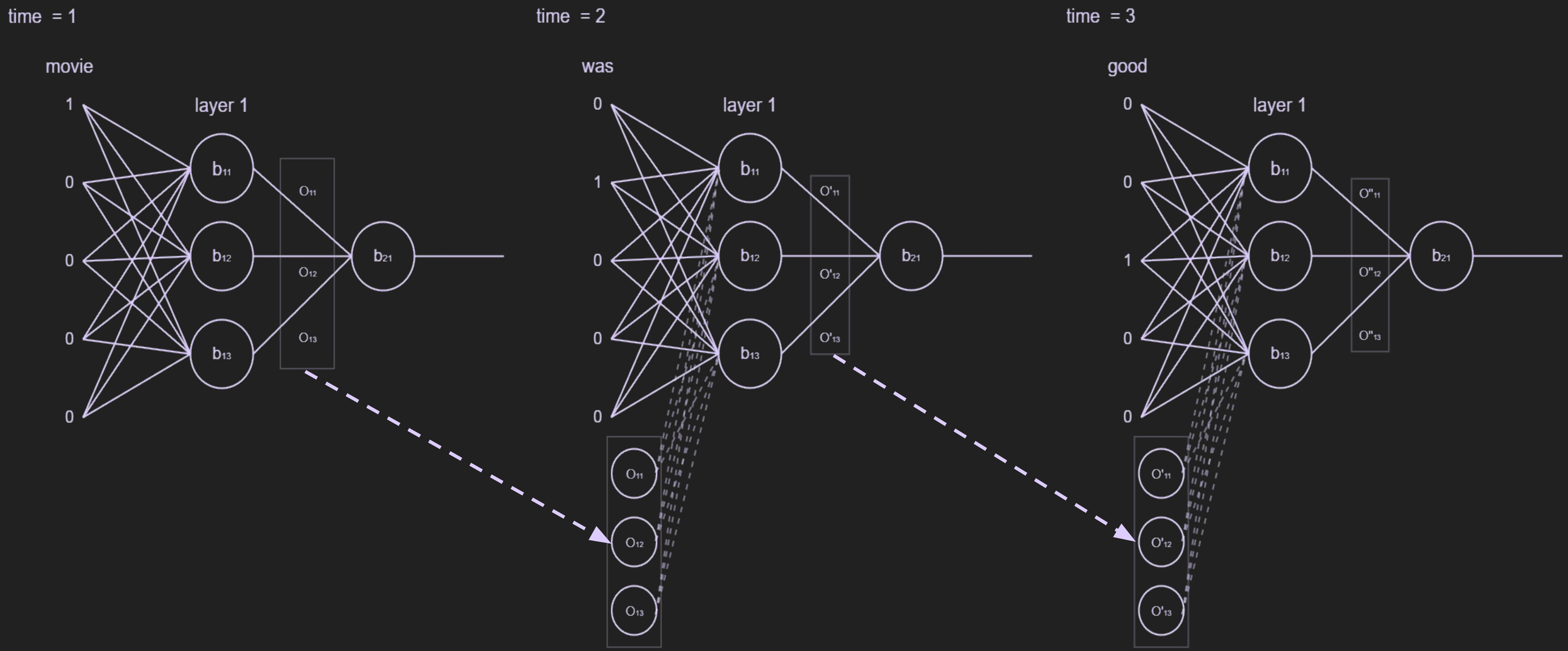
If we feed 1st word of 1st review then time step = 1

(1 , [1,0,0,0,0])

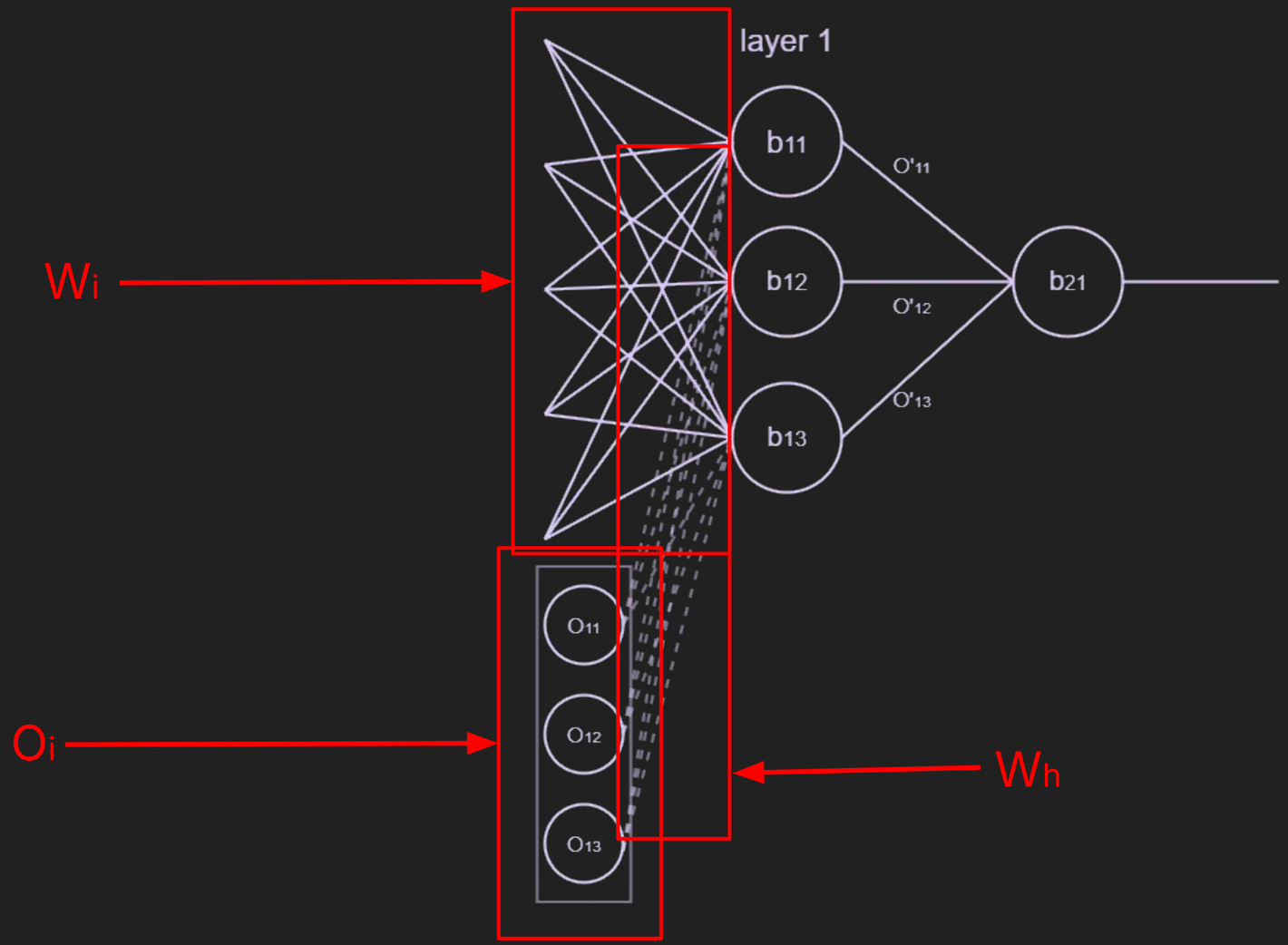
How RNN works ?

Let's say we are doing sentiment analysis on the movie review data and we have following architecture -




time = 1
movie
1
0
0
0
0
layer 1
b11
b12
b13
O11
O12
O13
b21
time = 2
was
0
1
0
0
0
layer 1
b11
b12
b13
O'11
O'12
O'13
b21
O11
O12
O13
time = 3
good
0
0
1
0
0
layer 1
b11
b12
b13
O''11
O''12
O''13
b21
O'11
O'12
O'13

How to train RNN ?

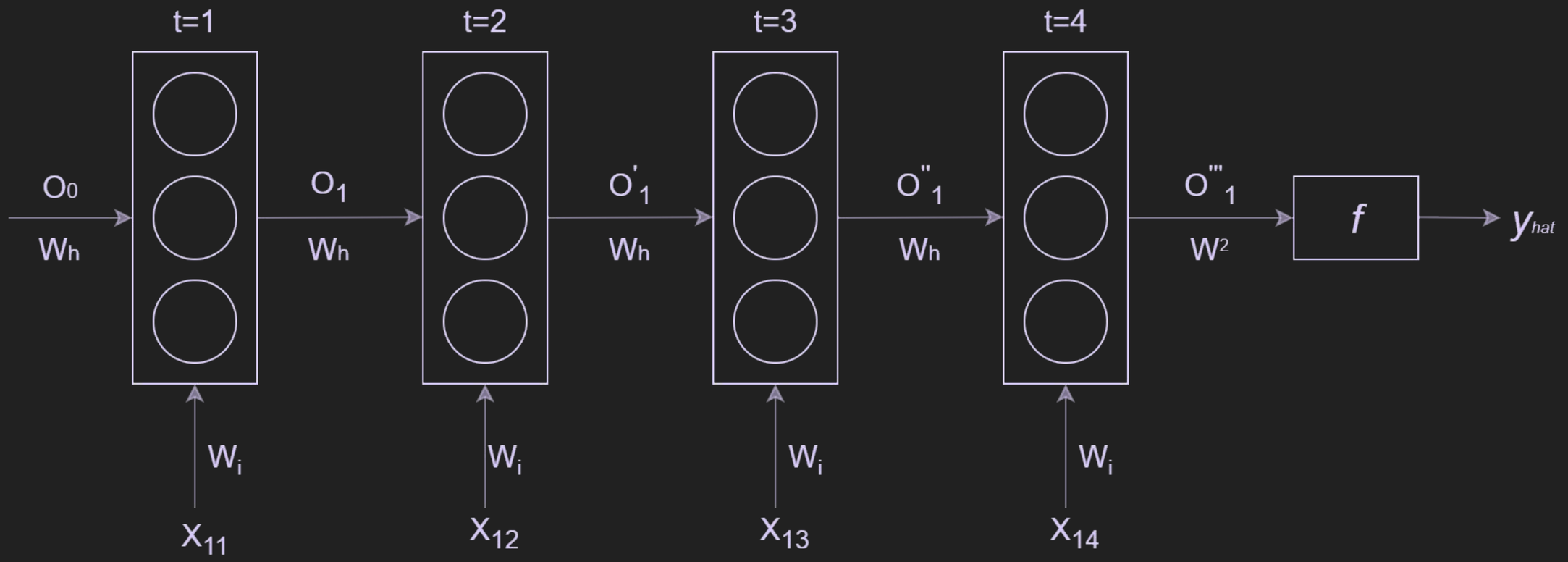


Step 0 - Initialise O_0 and W_h

Step 1 - Select a row from data and do forward propagation

Step 2 - Calculate loss

Step 3 - Use gradient descent to update weights and bias



(Here I assumed sentence have 4 words)

$$
\begin{aligned}
O_1 &= f(X_{11}W_i + O_0W_h + b_1) \\
O_1' &= f(X_{12}W_i + O_1W_h + b_1) \\
O_1'' &= f(X_{13}W_i + O_1'W_h + b_1) \\
O_1''' &= f(X_{14}W_i + O_1''W_h + b_1)
\end{aligned}
$$

$$
\hat{y} = f(O_1'''W^2 + b_{21})
$$

$$
\frac{dL}{d\hat{y}} \qquad \frac{dL}{dW^2} = \frac{dL}{d\hat{y}} \cdot \frac{d\hat{y}}{dW^2}
$$

$$
\frac{dL}{dW^i} = \frac{dL}{d\hat{y}} \cdot \frac{d\hat{y}}{dO_1'''} \cdot \frac{dO_1'''}{dW^i}
$$

$$
W^2_{new} = W^2 - \eta \cdot \frac{dL}{dW^2}
$$

$$
W_{i\,new} = W_i - \eta \cdot \frac{dL}{dW_i}
$$